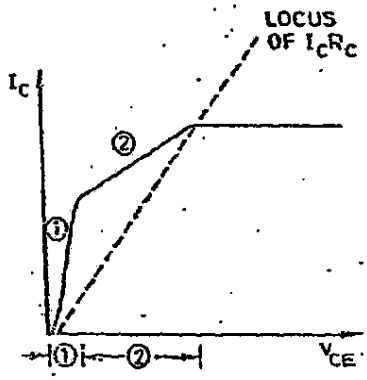
- The high-voltage power MOS has the advantage of high-frequency performance.
- Base transit time dominates the maximum $f_T$ of the power bipolar transistor when the device approaches high-level injection.

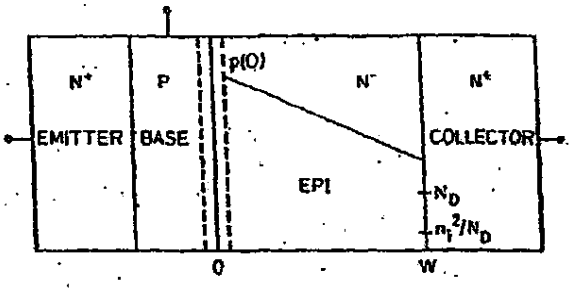
Another basic difference between MOS and bipolar devices (not related to Johnson's limit) is the ability of bipolar transistors to achieve very low on-resistance through "conductivity modulation." The saturation characteristics of high-voltage $N^+P\bar{N}\bar{N}^+$ bipolar transistors often exhibit two pronounced regions--the saturation ① and quasi-saturation ② regions as illustrated in Fig. 3.31 [3.55]. When the transistor is in saturation, the P-$\bar{N}$ base-collector junction is forward biased and the high-level injection of minority carriers (holes) from the base into the $\bar{N}$ layer modulates all or part of the conductivity of this layer, thereby reducing collector resistance. The carrier distributions for ① and ② are plotted in Fig. 3.32.



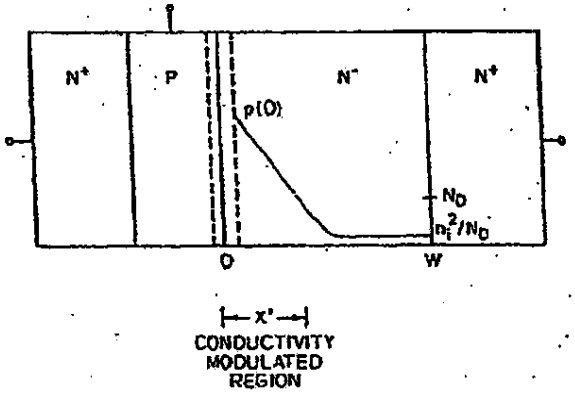

① = SATURATION
② = QUASI-SATURATION

Fig. 3.31. COMMON-EMITTER CHARACTERISTICS OF THE $N^+P\bar{N}\bar{N}^+$ BIPOLAR POWER TRANSISTOR WITH TWO REGIONS OF SATURATION.

In region ①, the transistor is so heavily saturated that the injected hole concentration is greater than the impurity concentration $N_D$ over the entire $\bar{N}$ layer. To maintain charge neutrality, the majority carriers (electrons) are supplied from the $N^+$ substrate. The electron concentration is approximately equal to the hole concentration and, as a result, $\bar{N}$ layer resistivity is lowered by the excess carriers. Under this condition and based on the assumption that the hole lifetime is sufficiently large to make the hole diffusion length much greater than the $\bar{N}$-region thickness, the




a. Region ① in Fig. 3.31




b. Region ② in Fig. 3.31

Fig. 3.32. CARRIER CONCENTRATIONS IN THE N⁻ LAYER.

voltage drop across the N⁻ layer $V_c$ can be calculated by solving one-dimensional current-transport equations, resulting in [3.56]

$$V_c \cong \frac{kT}{q} \ln\left[\frac{p(0)}{N_D}\right] \qquad (3.52)$$

where

k = Boltzmann's constant

T = temperature (°K)

p(0) = injected hole concentration at the base-collector junction

The magnitude of $V_c$ can be estimated for a power transistor having $p(0)/N_D$ = 300 (base doping/collector doping). Substituting into Eq. (3.52) yields $V_c \approx 150$ mV which indicates that the voltage drop across the $N^-$ layer is determined by the injection level, not by the absolute value of $N_D$.

This advantage would be lost, however, if the bipolar transistor is not as heavily saturated and the hole injection into the $N^-$ layer is not sufficient enough to conductivity modulate all of the layer (region ② in Figs. 3.31 and 3.32b). The base-collector junction is less forward biased as collector current is increased. The unmodulated resistance of the collector region will contribute to the total voltage drop in addition to the voltage drop across the modulated portion as defined in Eq. (3.52). Collector resistance increases as the base-collector junction is less forward biased, and the bipolar on-resistance approaches that of the MOS.

Because the number of excess minority carriers in the collector also depends on minority lifetime, a direct trade-off exists between on-resistance and switching speed. As minority lifetime becomes longer in the base and collector, on-resistance becomes smaller but at the expense of switching speed. Reducing lifetime will increase the speed, but it will also raise the on-resistance because it limits the extent of conductivity modulation (X' in Fig. 3.32b).

The one-dimensional structure in Fig. 3.33 is used to explain why the power MOSFET has a higher on-resistance. The two $N^+$ regions (source and drain) are long enough to maintain thermal equilibrium.

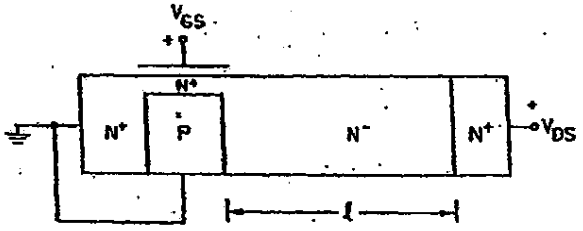



Fig. 3.33. ONE-DIMENSIONAL MOS USED TO EXPLAIN HIGH ON-RESISTANCE AND SINGLE INJECTION.

When $V_{DS}$ and $V_{GS}$ are applied with the polarities shown, the $N^+N^-$ junction is reverse biased and the surface $N^-$ channel is formed. The electrons are injected from the $N^+$ channel into the $N^-$ region. When the injected electron density is much less than the background doping density (low-level injection), the device is in the ohmic region and current flow is

$$J = \frac{q\mu_n N_D V_{DS}}{\ell} \qquad (3.53)$$

where

$N_D$ = doping concentration in the $N^-$ region

$\mu_n$ = electron bulk mobility

$\ell$ = length of the $N^-$ region

The voltage drops at the $N^+$ source and drain and the channel are normally negligible.

When injected electron density is comparable to background doping density, quasi-neutrality cannot be maintained because of the lack of available holes to compensate for the excess of electrons; the two $N^+$ regions have a very low concentration of holes ($\approx 2.1\ cm^{-3}$ for $N^+ = 10^{20}\ cm^{-3}$) and are incapable of injection into the $N^-$ region. The $PN^-$ junction is reverse biased and, as a result, the P-region is also incapable of injecting holes. At this point, injected charge becomes space charge and the current is completely space-charge limited (SCL). Based on the assumptions that current is carried chiefly by drift and that diffusion and the dependence of mobility on the electric field are not taken into account, SCL current can be expressed [3.57] as

$$J_{SCL} = \frac{9\varepsilon_{si}\mu_n}{8\ell^3} V_{DS}^2 \qquad (3.54)$$

The crossover voltage $V_{CO}$ at which the ohmic behavior changes to space-charge-limited behavior is obtained [3.58] by combining Eqs. (3.53) and (3.54),

$$V_{CO} = \frac{8qN_D\ell^2}{9\varepsilon_{si}} \qquad (3.55)$$

This crossover voltage can also be derived by equating the transit time of the injected carrier to the dielectric relaxation time of the $N^-$ region; that is,

$$\frac{\ell}{\mu_n V_{CO}/\ell} \cong \frac{\varepsilon_{si}}{\sigma} = \frac{\varepsilon_{si}}{N_D q \mu_n} \qquad (3.56)$$

or

$$V_{CO} \cong \frac{qN_D\ell^2}{\varepsilon_{si}} \qquad (3.57)$$

which is identical to Eq. (3.55) except for the numerical values. This concept can be understood by noting that, if dielectric relaxation time is larger than transit time ($V_{DS} > V_{CO}$), the injected carriers do not decay significantly while in transit through the $N^-$ region and SCL current is maintained. When transit time is greater than relaxation time ($V_{DS} < V_{CO}$), however, quasi-neutrality holds and the current becomes ohmic.

The $N^+N^-N^+$ diode at high-level injection is therefore considered to be "single injection" (electrons only), which is in direct contrast to the $P^+N^-N^+$ diode. When the $P^+N^-N^+$ diode is forward biased, holes are injected from the $P^+$ into the $N^-$ region and, to maintain charge neutrality, the electrons are injected from the $N^+$ to the $N^-$ region to "neutralize" the excessive holes. The $P^+N^-N^+$ diode thereby has a "double injection" (both electrons and holes). When excess carriers of only one type are injected, the current becomes SCL; however, if carriers of the other type are available to relax the injected-carrier space charge, considerably higher currents can be obtained. When the density of the injected carrier exceeds the background doping density, conductivity-modulation effects are observed.

At high-level injection, the voltage drop across the $N^-$ layer is considerably different in the $N^+N^-N^+$ and $P^+N^-N^+$ diodes. An estimate

of $V_{OC}$ in the $N^+N^-N^+$ diode can be obtained from Eq. (3.57). When $N_D = 3.5 \times 10^{14}$ cm$^{-3}$ and $L = 33.57$ μ, $V_{OC}$ is 604 V; in contrast, the voltage drop in the $N^-$ layer for the $P^+N^-N^+$ is ≈150 mV [Eq. (3.52)].

In the high-voltage VDMOS and under strong gate bias and low $V_{DS}$, the $N^-$ region below the gate becomes the $N^+$ accumulation layer and, therefore, the VDMOS can be considered an enhancement MOS transistor in series with an $N^+N^-N^+$ diode. At low $V_{DS}$, the device is ohmic (low-level injection in the $N^+N^-N^+$ region). As calculated above, high-level injection requires $V_{DS}$ to be ≈604 V. The enhancement MOS transistor will saturate long before this voltage is reached and, as a result, the space-charge-limited behavior is not observed in the power MOSFET.

In the high-voltage bipolar transistor, the base-collector region is essentially a $P^+N^-N^+$ diode. When the bipolar transistor is in heavy saturation, this diode is forward biased and is at high-level injection which results in a very low voltage drop as discussed above.

In conclusion, the vertical structures of the power MOS and bipolar devices are similar; however, their on-resistances differ greatly because of the fundamental differences in carrier injection and the transport mechanism. In the MOS, on-resistance is determined by the epitaxial bulk resistance. In bipolar devices, on-resistance is determined by the external biases and injection levels and epitaxial resistance is not important.

## F. Summary

This chapter has described the wide range of voltage limitations existing in high-voltage power MOS transistors, the parasitic bipolar effects, and the ultimate Johnson limit. The analysis of dielectric breakdown led to the conclusion that oxide rupture can only occur between the gate and source, not between the gate and drain. Among the junction-edge termination techniques, the selection of one depends on efficient utilization of the silicon area and processing compatibility. The analysis of the punchthrough and bipolar latchback limits indicated that the switching requirement dV/dt places a constraint on channel thickness and doping under certain layout conditions and before bipolar turn-on. The punchthrough limit on the channel profile is effective only when the

channel contacts are adjacent to the gates. Bipolar turn-on caused by weak avalanche in the static condition is very unlikely unless avalanche of the channel-drain junction is initiated near or under the gate; one such possibility is that the groove is etched too deeply in the VMOS.

Johnson's limit was considered for various devices. A new limit was defined, based on recent experimental bulk scattering-limited velocity data. The differences between MOS and bipolar transistors were described in terms of frequency response and on-resistance.

Other limits in power transistors include the maximum allowable channel temperature, thermal properties of the device and package, and thermal fatigue. The maximum power-handling capability may be limited by these factors.

# Chapter IV

## ELECTRON MOBILITY IN INVERSION AND ACCUMULATION LAYERS ON THERMALLY OXIDIZED SILICON SURFACES

Electron mobilities in surface inversion and accumulation layers are fundamental parameters in device design and circuit simulation. Their variations with substrate doping, vertical electric field, crystal orientation, substrate or back gate bias, temperature, and fixed oxide charge $Q_f$ must be known quantitatively, therefore, if MOS transistor models are to simulate the actual devices. It has become apparent in modeling high-performance power MOSFETs (DMOS and VMOS) [4.1] that sufficient data for accurate modeling are not available in the literature.

Earlier experimental work on inversion-layer mobility has concentrated on Hall and field-effect mobilities [4.2,4.3,4.4]. There is very little existing data concerning the effective electron mobility in surface inversion layers (also known as conductivity mobility). This effective mobility appears in all theoretical models of MOS transistors and, as a result, its measurement is essential in MOS device modeling.

In most MOS structures, majority-carrier surface mobility plays an important role in device performance; depletion-mode transistors commonly employed in NMOS integrated circuits and the $N^-$ drift region in the DMOS and power MOS are two examples. Electron mobility in surface accumulation layers has been characterized but only for a single-crystal orientation and one substrate-doping level [4.5].

High gate voltages reduce carrier mobility in the surface channel below it, and this factor must be considered in the modeling of MOS I-V characteristics as a function of gate drive. In addition, the knowledge of mobility variations with oxide charge and substrate doping concentration is essential for the accurate calculation of on-resistance, transconductance, and drain current. For example, the experimental mobility data described in this chapter have been used in several key equations in Chapter II, such as enhancement-mode channel resistance $\mu_E(V_G)$ in Eq. (2.10), depletion-mode on-resistance $\mu_D(V_G)$ in Eq. (2.13), device transconductance $g_E$ in Eq. (2.[illegible]), and enhancement-mode drain current $I_D$ in Eq. (2.50).

In this chapter, empirical relationships for mobility are derived as a function of the relevant substrate and electrical properties. These equations are valid over a wide range of parameters, which implies that they should be useful in device modeling and circuit simulation. Based on the observed variations of high-field mobility with the processing parameters, several conclusions concerning the optimal process conditions for maximizing mobility will be reached.

## A. Experimental Techniques

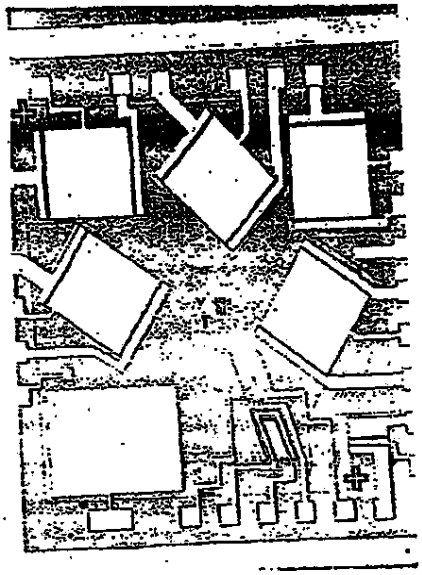
The test structures used were large (500 × 500 $\mu m^2$) MOS transistors for the surface-mobility measurements and MOS capacitors for the measurement of oxide thickness, average surface doping concentration, and interface charge density ($Q_f$ and $N_{it}$). The devices were fabricated simultaneously on (100), (110), and (111) boron-doped substrates with concentrations ranging from $3 \times 10^{14}$ to $1.7 \times 10^{17}$ $cm^{-3}$. Five MOS transistors with current flow in the major surface directions with respect to the three crystal orientations were included. Figure 4.1a is a photomicrograph of the test chip, and Fig. 4.1b shows the crystal orientations of the substrates and directions of current flow. The fixed oxide charge density $Q_f$ was varied on separate wafers by using different oxidation temperatures (700° to 1100°C) and fast pulls from an oxygen ambient. Minimum $Q_f$ values were obtained after $N_2$ anneals following oxidation. A metal-gate process was selected for device fabrication.

Effective mobility was derived from dc drain-conductance $g_d$ measurements by means of a calculator-controlled automatic system [4.6] and was found to be related to drain conductance by

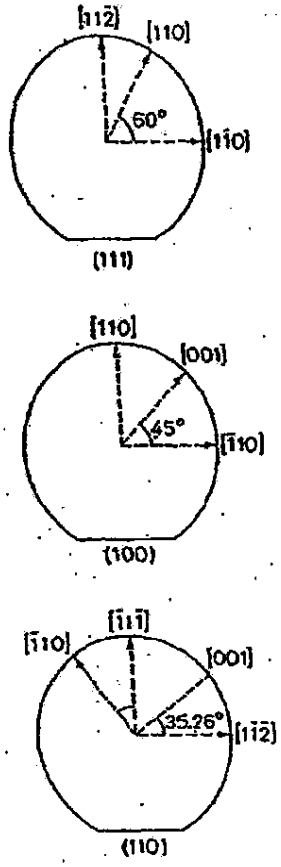
$$\mu_{eff} = \left. \frac{(L/W)\, g_d}{Q'_{inv}} \right|_{V_D \to 0} \tag{4.1}$$

where the total induced charge in the channel per unit area [4.7] is

$$Q'_{inv} = C_o \left[ V_G - V_T - \frac{1}{2} \left( 1 + \frac{a}{2\sqrt{\beta \phi_{so}}} \right) V_D \right] \tag{4.2}$$



a. Photomicrograph




b. Crystal orientations of substrates and directions of current flow

Fig. 4.1. EXPERIMENTAL TEST DEVICE FOR MOBILITY MEASUREMENTS.

Here, $C_o$ is the measured gate capacitance per unit area, $\beta\phi_{so} \cong 2 \ln (N_A/n_i)$ is the band bending at the source end of the channel, $a = \sqrt{2}\,(K_s t_{ox}/K_{ox} L_B)$, and $L_B = \sqrt{kTK_s\varepsilon_o/q^2N_A}$ is the bulk Debye length. By extrapolation of the tangent at the point of inflection on the $I_D$ vs $V_G$ curve, the threshold voltage was determined from the zero current intercept $V_{Gi}$ to be

$$V_T = V_{Gi} - \frac{1}{2}\left(1 + \frac{a}{2\sqrt{\beta_{sc}}}\right)V_D \qquad (4.3)$$

It should be noted that effective mobility is distinct from the field-effect mobility $\mu_{FE}$ obtained from transconductance $g_m$ and defined as

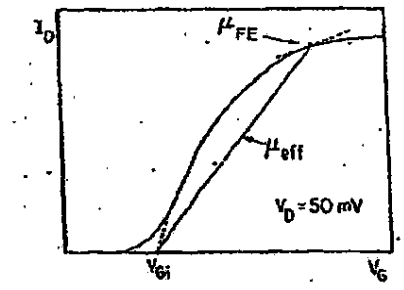
$$\mu_{FE} = \frac{L/W}{C_o V_D}\left(\frac{dI_D}{dV_G}\right)\Bigg|_{V_D \to 0} \qquad (4.4)$$

The relationship between $\mu_{eff}$ and $\mu_{FE}$ is plotted in Fig. 4.2a where it can be seen that the mobility data obtained from the $g_d$ and $g_m$ measurements are identical at the point of maximum slope (near $V_T$) for low fast surface state density. Beyond the point of inflection, however, $\mu_{FE} < \mu_{eff}$ as is demonstrated in Fig. 4.2b. If $\mu_{FE}$ is used in device modeling, therefore, the currents and switching speeds will be underestimated.

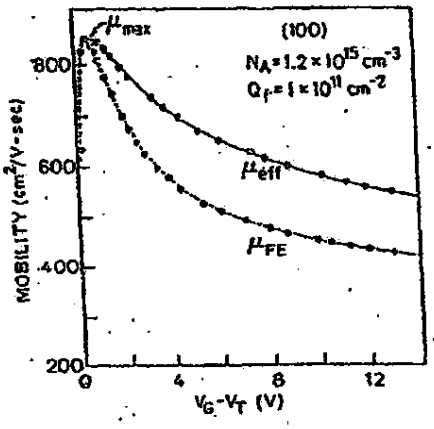
B. Physical Mechanisms in Surface Carrier Scattering

Early analyses of electron mobility in surface inversion layers were generally based on the classical theory of diffused or partially diffused surface scattering [4.8,4.9,4.10]. In recent years, there have been extensive studies of the different scattering mechanisms in terms of the interface properties [4.11-4.14]. This section briefly reviews some of these investigations to serve as a basis for the discussion of the experimental results presented in this chapter.

Figure 4.3 is a schematic cross section of the $Si/SiO_2$ interface region of a MOSFET and its charge distributions; the irregularity of the interface is also shown. This structure has been discussed by Cheng and Sullivan [4.11], and a similar model based on Auger studies has been proposed [4.15]. The four types of charges existing at or near the interface are positive fixed oxide charge $Q_f$, interface state charge $N_{it}$ which can be either positive or negative depending on whether it is a donor or acceptor, bulk ionized impurity charge $Q_B$, and induced mobile electrons whose distribution is shown as a function of distance from the surface.




(a)




(b)

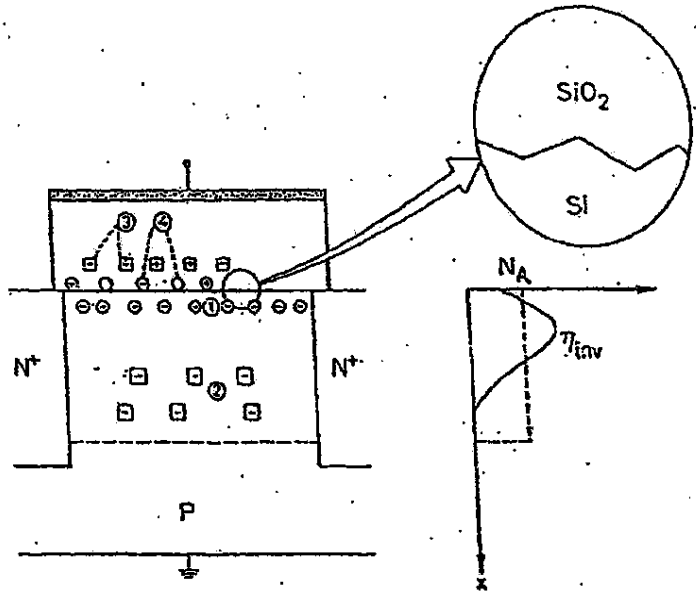
Fig. 4.2. DEFINITION AND RELATIONSHIP BETWEEN EFFECTIVE MOBILITY $\mu_{eff}$ AND FIELD-EFFECT MOBILITY $\mu_{FE}$.

The following scattering mechanisms may account for the mobility behavior in the region where the gate voltage is above threshold.

- Phonon scattering caused by the various modes of lattice vibration, including surface acoustic and optical phonons [4.13]: this scattering is significant at room temperature but is negligible at very low temperatures.

- Coulomb scattering as the result of charged centers, including fixed oxide charge, interface state charge, and localized charge caused by ionized impurities [4.12,4.13]: its effects are important in lightly inverted surfaces; high surface charge densities or substrate doping concentrations imply increased coulomb scattering. It becomes less pronounced in a heavily inverted surface because of carrier screening.

- Surface-roughness scattering resulting from deviation of the interface from an ideal plane [4.11,4.14]: this type of scattering is dominant under strong inversion because the strength of the interaction is governed by the distance of the carriers from the surface; the closer the carriers are to the surface, the stronger the scattering.




① inversion-layer electron charge
② ionized impurity charge
③ fixed oxide charge $Q_f$
④ interface state charge $N_{it}$

Fig. 4.3. THE Si/SiO2 INTERFACE AND ASSOCIATED ELECTRICAL CHARGES IN A MOSFET STRUCTURE.

The effectiveness of these mechanisms depends on the operating temperature and strength of the surface electric field. At low temperatures, mobility is determined by coulomb scattering which dominates the low gate field and by surface-roughness scattering which dominates the high field. At room temperature, mobility is governed by coulomb scattering because of the charged centers and by phonon scattering in the low field. It is controlled by surface-roughness and phonon scattering under strong inversion.

In the weak-inversion (subthreshold) region, minority-carrier density fluctuations, generated by interface states and fixed oxide charges, are believed to account for the mobility variations [4.16]. At higher carrier densities (larger than $5 \times 10^{11}$ $cm^{-2}$), however, such fluctuations should have a negligible effect. Sections C and D will discuss mobility behavior only above threshold and, as a result, the observed data can be explained in terms of the three scattering mechanisms.

C. Electron Inversion-Layer Mobility--Effects of Oxide Charges and Substrate Resistivity on Maximum Effective Mobility

Mobility near the threshold voltage at room temperature is governed by scattering caused by the interface charged centers and phonons except in more heavily doped samples where surface-roughness scattering may also become important. The effect of the impurity concentration on maximum effective mobility $\mu_{max}$ is through the gate field required to produce surface inversion because mobility is a function of gate field rather than carrier density; more lightly doped samples have lower surface fields at the onset of inversion and correspondingly higher $\mu_{max}$.

Figure 4.4 plots $\mu_{max}$ at room temperature as a function of surface oxide charge density $Q_f$ at a number of substrate impurity concentrations. Mobility decreases hyperbolically with increasing $Q_f$ at a given doping level and is independent of surface orientation because both (111)- and (100)-oriented wafers were used. This behavior does not change with the doping concentration although, as expected, mobility does decrease as the concentration is increased. (It should be reemphasized that this reduction is not a result of impurity scattering but of the higher vertical fields required to produce surface inversion.)

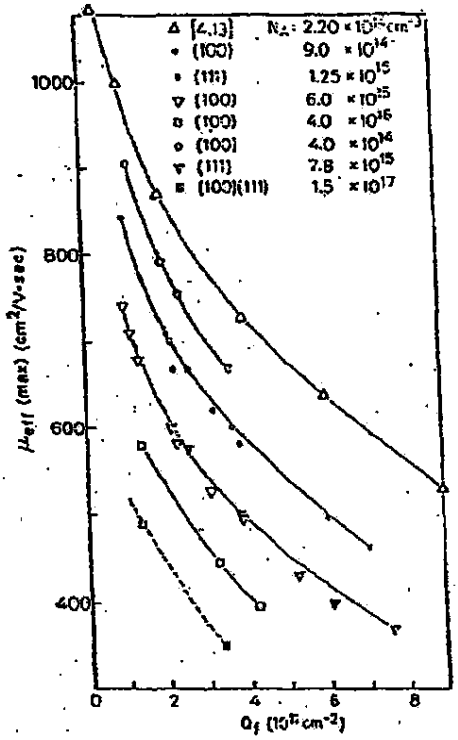



Fig. 4.4. PEAK EFFECTIVE MOBILITY AT ROOM TEMPERATURE AS A FUNCTION OF SUBSTRATE DOPING CONCENTRATION AND FIXED OXIDE CHARGE DENSITY.

The hyperbolic form of this mobility variation with $Q_f$ can be predicted theoretically from a calculation of scattering by charged centers where the relaxation time is inversely proportional to $Q_f$ and the relaxation time resulting from phonon and surface-roughness scattering is independent of $Q_f$ [4.17]. An empirical relationship for $\mu_{max}$ as a function of $Q_f$ and $N_A$ is defined, therefore, as

$$\mu_{max} = \frac{\mu_0(N_A)}{1 + \gamma(N_A)\, Q_f} \qquad (4.5)$$

where

$N_A$ = average surface impurity concentration ($cm^{-3}$)

$Q_f$ = oxide charge density in units of $10^{11}$ $cm^{-2}$

and $\alpha$ is a factor dependent on $N_A$ which must be determined experimentally.

To obtain the functional dependence of $\mu_o$ and $\alpha$ on $N_A$, the inverse of $\mu_{max}$ is plotted in Fig. 4.5 as a function of $Q_f$. The slope of each curve is proportional to $\alpha$ and the intercept is $\mu_o$. Based on this figure,

$$\mu_o = 3490 - 164 \log N_A$$

$$\alpha = -1.04 \times 10^{-1} + 1.93 \times 10^{-2} \log N_A$$

The accuracy of $\mu_{max}$ calculated from Eq. (4.5) is within ±4 percent of the experimental data, with $N_A$ ranging from $2 \times 10^{14}$ to $1.7 \times 10^{17}$ $cm^{-3}$ and $Q_f$ from $0.4 \times 10^{11}$ to $7 \times 10^{11}$ $cm^{-2}$. Included in the figure are data taken from the literature [4.13,4.18] which are also in agreement with the formalism expressed in Eq. (4.5).

It should be noted that the horizontal axis in Fig. 4.5 and the charge density in Eq. (4.5) are both termed $Q_f$. In most MOSFETs, the densities of other types of charged centers near the $Si/SiO_2$ interface are much smaller than $Q_f$. The densities of interface states $N_{it}$, for example, are generally an order of

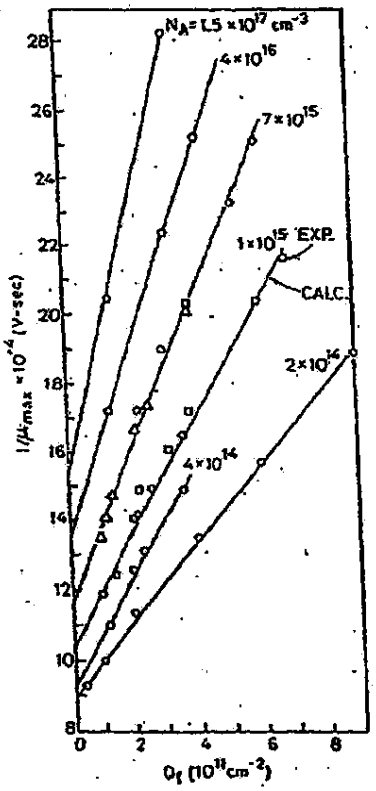



Fig. 4.5. INVERSE OF PEAK EFFECTIVE MOBILITY VS $Q_f$ AT VARIOUS SUBSTRATE DOPING CONCENTRATIONS.

magnitude smaller than $Q_f$ if an appropriate thermal anneal (400° to 500°C in an $N_2$ or $N_2/H_2$ ambient) is included in the fabrication sequence [4.19]; all data reported here were obtained after such an anneal. Additional experiments varied the ratio of $Q_f$ to $N_{it}$ while holding the sum of the two constant, and these experiments indicated that, when $N_{it}$ is not less than $Q_f$, the appropriate charge density to use in Eq. (4.5) is the sum of the two.

In addition, charges in the bulk of the oxide greater than 50 Å from the $Si/SiO_2$ interface have little effect on mobility [4.20], as was confirmed in this study by mobility measurements of samples with deposited $SiO_2$ dielectrics. Such CVD oxides can have substantial bulk charges [4.21]. For example, the mobility of a sample with a measured effective $Q_f$ of $\approx 7.6 \times 10^{11}/cm^2$ was found to be typical of a thermal oxide with $Q_f \approx 3.2 \times 10^{11}/cm^2$ (the surface mobility of the deposited oxide was substantially higher than would be inferred from its measured effective $Q_f$). The reason for this behavior was most likely because much of the effective $Q_f$ was bulk oxide charge located far enough from the $Si/SiO_2$ interface so as not to degrade mobility.

The empirical relationship expressed in Eq. (4.5) can yield an estimate of the mobility variation under various processing conditions and also indicates that deviations in the reported mobility data may be the result of variations in the density of interface charged centers. It should also be noted that the commonly used approximation wherein the surface inversion-layer mobility is assumed to be 50 percent of bulk mobility [4.22] is substantially in error under many conditions. For example, with reasonably low $Q_f$ densities ($1 \times 10^{11}$ $cm^2$), electron surface mobility $\mu_{max}$ varies between 75 and 85 percent of the bulk value at doping levels between $2 \times 10^{14}$ and $1 \times 10^{17}$ $cm^3$. Figure 4.6 compares the inversion-layer and bulk mobilities vs substrate doping level at two values of $Q_f$. It is apparent that $\mu_{max} \approx 1/2\ \mu_{bulk}$ is a closer approximation at higher $Q_f$; however, these values are not representative of MOS devices.

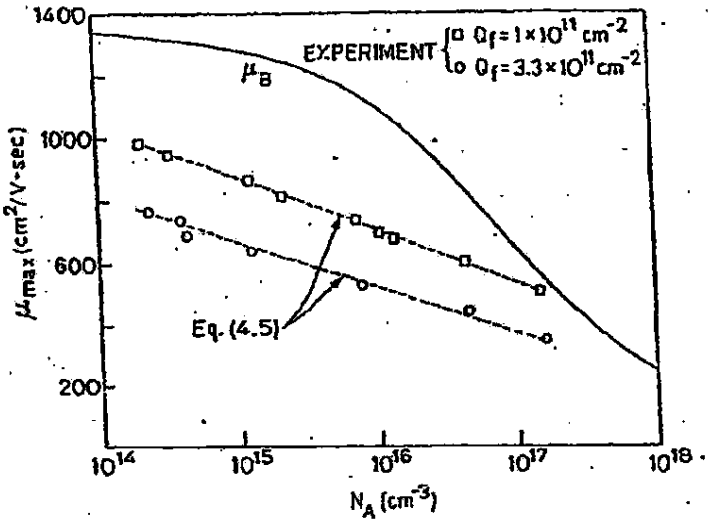



Fig. 4.6. INVERSION-LAYER PEAK MOBILITY $\mu_{max}$ AND BULK MOBILITY VS SUBSTRATE DOPING CONCENTRATION AT TWO VALUES OF OXIDE CHARGE DENSITY $Q_f$.

D. Electron Inversion-Layer Mobility--Variations of $\mu_{eff}$ with Vertical Electric Field

1. Peak Field at the Silicon Surface vs Avarage Field in the Inversion Layer

From Gauss's law, the peak field at the silicon surface is

$$E_{si} = \frac{1}{\varepsilon_{si}} (Q_{inv} + Q_B) \qquad (4.6)$$

where $Q_{inv} = qN_{inv}$ is the contribution from the induced mobile electrons in the inversion layer as defined in Eq. (4.2) and $Q_B$ is the depletion charge per unit area. This peak field, however, is not encountered by most of the electrons in the inversion layer. From quantum-mechanical calculations of inversion-layer carrier distributions [4.23], the carriers are distributed in a gaussian profile with the peak inside the silicon surface. If two substrate impurity doping levels are used

(even if $E_{si}$ is adjusted to be the same), the average field inside the two inversion layers can differ substantially, therefore, as illustrated in Fig. 4.7a. Two very different doping levels ($N_A = 1 \times 10^{17}$ and $1 \times 10^{14}$ $cm^{-3}$) were chosen to emphasize the difference in $Q_B$, and this difference causes the variations in the fields at the interface between the inversion and depletion layers. Consequently, the average field in the inversion layers $E_{eff}$ will differ even when the same $E_{si}$ is maintained. It would appear most appropriate, therefore, to consider mobility variation with $E_{eff}$[†] rather than with the more commonly used $E_{si}$.

The effective field in the inversion layer [4.24,4.25]

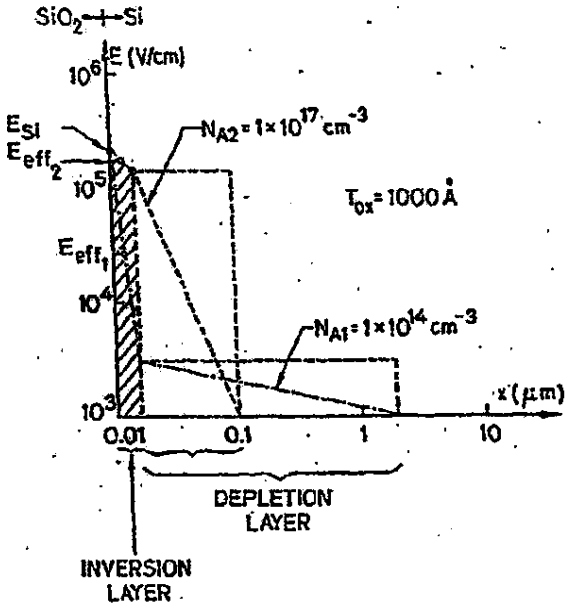
$$E_{eff} = \frac{1}{\epsilon_{si}} \left( \frac{1}{2} Q_{inv} + Q_B \right) \tag{4.7}$$

is illustrated in Fig. 4.7a, and the ratio of $E_{eff}$ values at two doping levels is plotted in Fig. 4.7b as a function of $E_{si}$. The lightly doped sample has a lower $E_{eff}$ for the same $E_{si}$, and the difference in $E_{eff}$ becomes smaller as $E_{si}$ increases. This partially explains why the published mobility curves for different doping levels tend to converge at high gate fields even if $E_{si}$ is used to plot the data.

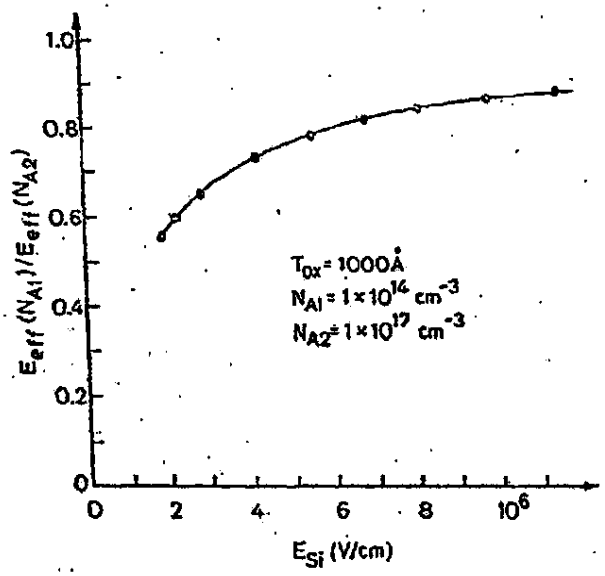
2. Mobility Variation with Effective Field

The concentration of bulk acceptors in or near the inversion layer is generally small; for example, $10^{15}$ $cm^{-3}$ bulk acceptors in a 100 Å inversion layer leads to $10^{9}$ ions/$cm^2$ which is less than the density of the oxide and interface state charges. When mobility is plotted as a function of $E_{eff}$, therefore, the effect of substrate impurity-doping levels can be largely removed at levels below $10^{17}$ $cm^{-3}$; mobility is determined primarily by the effective field [4.25] because coulomb scattering is the result of interface charges rather than ionized impurity charges. In Fig. 4.8, substrate doping ranges from $3 \times 10^{14}$ to $1.4 \times 10^{17}$ $cm^{-3}$. Except for the region near the onset of inversion in each

[†] Proposed by J. Clemens of Bell Laboratories.




a. Peak electric field at the silicon surface and average electric field in the inversion layer




b. Ratio of the two effective fields

Fig. 4.7. EFFECT OF SUBSTRATE DOPING CONCENTRATION ON THE AVERAGE ELECTRIC FIELD.

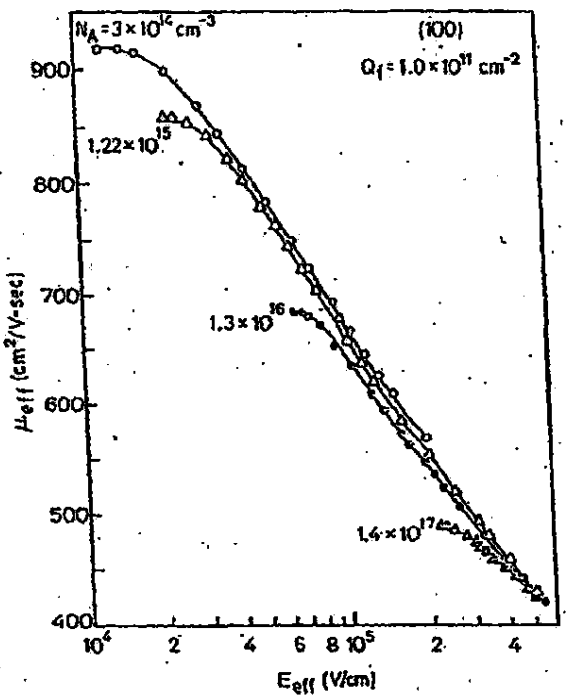



Fig. 4.8. EFFECTIVE MOBILITY VS EFFECTIVE FIELD FOR FOUR SUBSTRATE DOPING CONCENTRATIONS.

sample, the difference in the mobility values is within 5 percent in the strong-inversion region for a given field. It should be noted that $Q_f = 1 \times 10^{11}$ cm$^{-2}$ in all samples and that $Q_f$ can raise or lower the mobility curve.

The dependence of mobility on the gate field can be described by an empirical approximation in the form [4.26] of

$$\mu_{eff} = \mu_{max} \left( \frac{E_c}{E_{eff}} \right)^{C_1} \qquad (4.8)$$

where $\mu_{max}$ is the maximum value for a given doping level and $Q_f$ value, and $C_1$ is an empirical constant independent of $Q_f$ and a weak function of substrate doping.

$$C_1 = 0.341 - 5.44 \times 10^{-3} \log N_A \quad \text{(wet } O_2\text{)} \qquad (4.9a)$$

$$= 0.313 - 6.05 \times 10^{-3} \log N_A \quad \text{(dry } O_2\text{)} \qquad (4.9b)$$

Typically, $C_1$ is on the order of 0.26 for a steam oxide and 0.22 for a dry $O_2$ oxide. When normalized mobility $\mu_{eff}/\mu_{max}$ is plotted as a function of $E_{eff}$ on a log-log scale, $C_1$ is proportional to the slope of the curve and is a measure of the rate of mobility degradation with vertical field. The difference in $C_1$ between wet and dry $O_2$ grown oxides will be discussed in Section 4.

The constant $E_c$ depends on both $N_A$ and $Q_f$. In a simplified approach, however, $N_A$ is not taken into account because of the use of the effective field, and $E_c$ is expressed as a function of $Q_f$ only,

$$E_c = Ae^{BQ_f} \qquad (4.10)$$

where, for wet oxidation, $A = 2.79 \times 10^4$ V/cm and $B = 8.96 \times 10^{-2}$ and, for dry oxidation, $A = 2.61 \times 10^4$ V/cm and $B = 0.13$ (both for $N_A = 4 \times 10^{14}$ cm$^{-3}$); $Q_f$ is in units of $10^{11}$ cm$^{-2}$. The reason for using a lightly doped sample is to extend the range of the effective field where the mobility can be represented. By inserting Eqs. (4.9) and (4.10) into Eq. (4.8), mobility can be calculated as a function of gate field for various $Q_f$ values. It should be emphasized again that the effect of substrate doping has been incorporated into the calculation of $E_{eff}$.

Past deficiencies in using one mobility curve for a lightly doped sample to represent a wide range of substrate resistivities have centered around the large discrepancy between the mobility of a lightly doped sample for a given field and the mobility of a heavily doped sample when the corresponding field is near its threshold voltage. This problem, clearly seen in Fig. 4.8, can be minimized by incorporating the $N_A$ dependence in Eq. (4.10) which yields

$$E_c = E_{co}(N_A)\, Ae^{BQ_f} \qquad (4.11)$$

The dependence of $E_{co}$ on $N_A$ is plotted in Fig. 4.9 and, based on a straight-line approximation to this data,

$$E_{co}(N_A) = 2.054 \times 10^{-4} N_A^{0.25} \qquad (4.12)$$

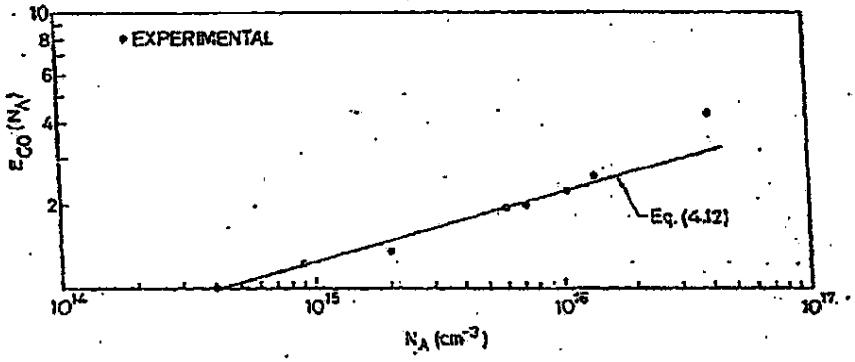



Fig. 4.9. DEPENDENCE OF THE CONSTANT $E_{co}$ ON DOPING CONCENTRATION $N_A$.

Although the influence of substrate doping on the mobility variation with field is small, except at very high concentrations and low fields, oxide charge density $Q_f$ has a strong impact on mobility over a wide range of surface fields. As can be seen in Fig. 4.10, the effect of $Q_f$ is most pronounced at low fields; it decreases at high fields but is still very important over most regions of device operation.

3. Mobility Variation with Substrate Bias

Applying substrate bias $V_{BS}$ changes the bulk depletion charge $Q_B$ ($Q_B$ is proportional to $(2\phi_f + V_{BS})^{1/2}$) and increases the threshold voltage which can now be expressed as

$$V_T(V_{BS}) = V_{FB} + 2\phi_F + \frac{Q_B(V_{BS})}{C_o} \qquad (4.13)$$

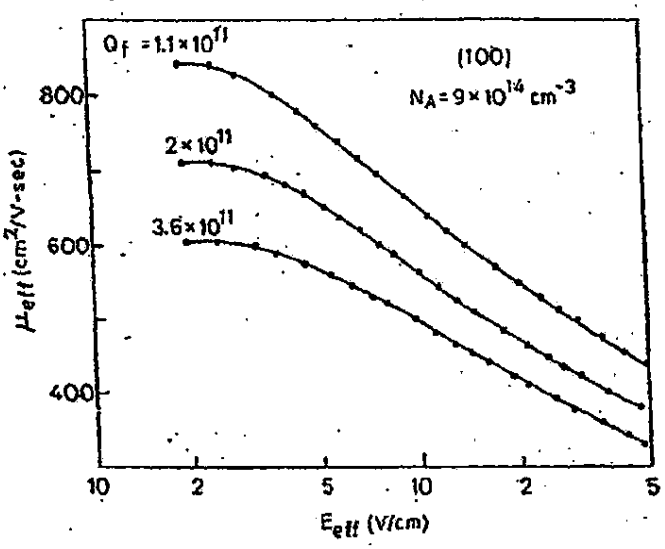



Fig. 4.10. EXPERIMENTAL ELECTRON INVERSION-LAYER MOBILITY VS EFFECTIVE FIELD AT VARIOUS VALUES OF $Q_f$.

Calculation of the effective field in the inversion layer is the same as in Eq. (4.7) except for the new values of $V_T$ and $Q_B$; $E_{eff}(V_{BS})$ now becomes

$$E_{eff}(V_{BS}) = \frac{C_o}{\epsilon_{si}}\left\{\frac{1}{2}\left[V_G - V_T(V_{BS})\right] + \frac{Q_B(V_{BS})}{C_o}\right\} \qquad (4.14)$$

The effect of $V_{BS}$ on mobility is to increase the surface field at the onset of inversion because of the rise in bulk charge. This is equivalent to a larger substrate impurity concentration but with no variation in the impurity concentration inside the inversion layer and, if the effective (average) gate field is used, the mobility curve of $V_{BS} = 0$ is still applicable [4.25]. The experimental data obtained at two doping levels with $V_{BS} = -2$ and $-4$ V are plotted in Fig. 4.11. Except near the threshold voltage where there is a small difference, the mobility curve of $V_{BS} = 0$ is well represented at various substrate biases.

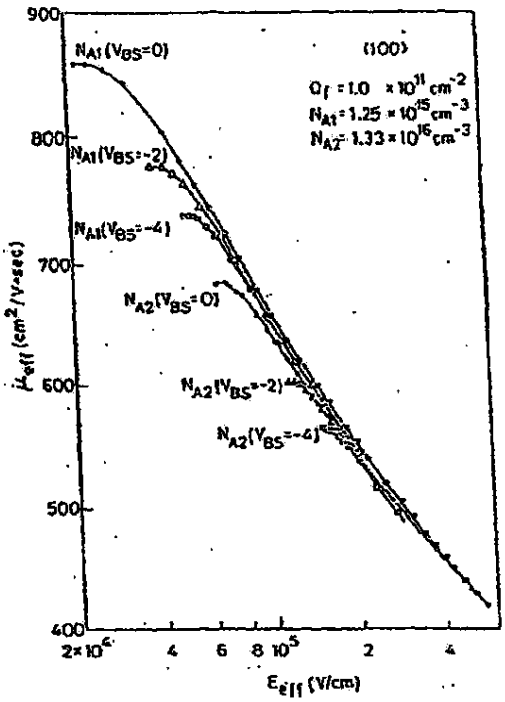



Fig. 4.11. EXPERIMENTAL ELECTRON INVERSION-LAYER MOBILITY VS EFFECTIVE FIELD AT TWO SUBSTRATE DOPING LEVELS. $V_{BS}$ = 0, -2, and -4 V.

4. Mobility Variation at High Fields with Dry vs Wet $O_2$ Thermal Oxidation

The effect of gate-oxidation conditions on carrier mobilities has been studied, using (100) p-type substrates; two types of gate oxides were grown in 950°C wet (95°C $H_2O$) and 1100°C dry $O_2$. After the $O_2$ cycle, low $Q_f$ samples were annealed in $N_2$ for 15 min at the oxidation temperature. Oxide thickness was ≈1000 Å. The normalized mobility measurements of three samples with the same substrate impurity concentration are plotted in Fig. 4.12 where it can be seen that the slope of the mobility curve of the dry-oxide sample with $Q_f = 1.4 \times 10^{11}\ cm^{-2}$ differs from the two wet-oxide samples; instead of being bracketed by the other curves, it drops more slowly with gate field. The dominant scattering mechanisms involved at high fields may help to interpret these differences.

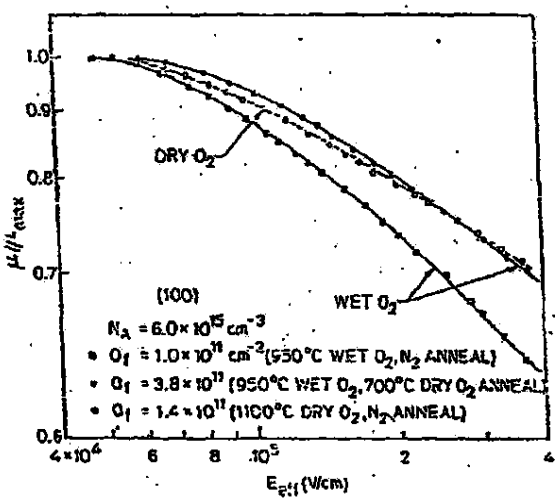



Fig. 4.12. NORMALIZED ELECTRON INVERSION-LAYER MOBILITY VS EFFECTIVE FIELD FOR DRY AND WET $O_2$ (95°C $H_2O$) SAMPLES.

It has been demonstrated [4.11] that the dominant mechanism is surface-roughness scattering under strong inversion; the closer the carriers are to the surface, the stronger the scattering. The oxidation process may vary the degree of smoothness of the silicon surface. It is well known that $SiO_2$ has a lower growth rate in a dry $O_2$ ambient, which may allow the interface more time to arrange its structure with less geometrical disorder, and a smoother surface may result. Similar behavior has been observed in samples with various doping levels. These results are not sensitive to the impurity concentration but only to the average electric field in the inversion layer. The present data thus indicate that a dry $O_2$ ambient results in higher electron mobility under strong inversion which will be even more pronounced at lower operating temperatures.

The presence of HCl in the oxidation ambient has also been studied. A gate oxide was grown at 1000°C in a mixture of 2 percent HCl and dry $O_2$. Small amounts of HCl in oxygen increase the thermal-oxidation rate of silicon relative to that of dry oxygen [4.27]; however, this increase is very small in comparison to that of a wet $O_2$ ambient. The

measurements indicate that there is little difference in the mobility rolloff between dry $O_2$ and HCl-$O_2$ oxides at high fields, which is in accordance with the assertion that surface roughness may be related to the oxidation rate.

It is interesting to note that a high-temperature $N_2$ anneal of one of the wet $O_2$ samples did not change the shape of the high-field mobility rolloff although it did reduce $Q_f$. It can be concluded, therefore, that the degree of surface roughness was not altered by the $N_2$ anneal even though the fixed oxide charge density was reduced substantially.

### 5. Mobility Variation at High Fields with Crystal Orientation and Surface-Current Direction

To investigate the effects of wafer orientation on high-field mobility, (100) and (111) samples with similar substrate resistivities and oxide charge densities were fabricated so as to avoid any possible confusion in interpreting the experimental data. The normalized mobilities of (111) and (100) orientations with $Q_f = 2.2 \times 10^{11}$ cm$^{-2}$ and $N_A = 6.5 \times 10^{15}$ cm$^{-3}$ are plotted in Fig. 4.13. In the high-field region, (100) mobility rolls off more slowly than does (111) mobility, and scattering caused by surface irregularities can still have a strong influence

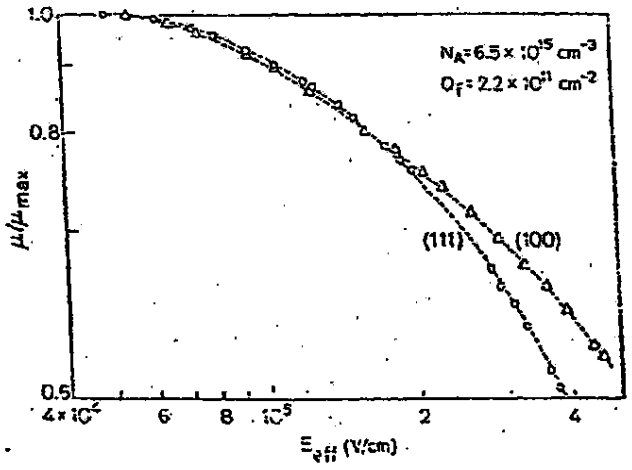
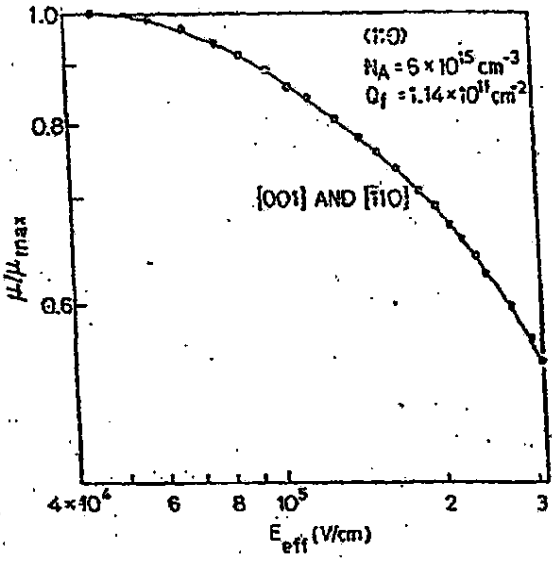



Fig. 4.13. NORMALIZED ELECTRON INVERSION-LAYER MOBILITY VS EFFECTIVE FIELD FOR (100)- AND (111)-ORIENTED WAFERS.

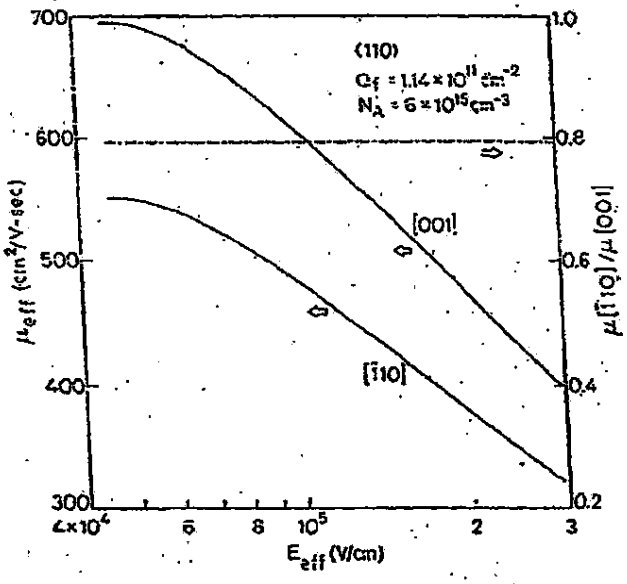
at room temperature and high fields even though phonon scattering cannot be neglected. This room-temperature data may indicate that the (100) samples are smoother than the (111) samples, which agrees well with a similar observation at low temperature (4.2°K) and high gate voltage (24 V) [4.28]. It would be difficult to compare two sets of wafers with different orientations because surface preparation (such as the degree of mechanical and chemical polishing) may vary; nevertheless, the behavior in Fig. 4.13 has been observed over a wide range of substrate resistivities, which is a strong indication that the mobility behavior of the (100) samples at high fields is different from that of the (111) wafers. The experimental data obtained in this study are contrary to an earlier postulate that the effect of scattering may be less for a (111) surface than for a (100) surface as the result of effective-mass considerations [4.2]. Because the (100) surface has a lower oxide growth rate and a similar high-field behavior was observed in Fig. 4.12 for slow vs fast oxidation, these results imply that the interface irregularities between Si and $SiO_2$ may be a function of the oxide growth rate. A slower oxidation process tends to produce slower mobility rolloff at high gate fields.

Earlier work [4.29,4.30] has demonstrated electron mobility anisotropy on (110) silicon wafers for a given surface orientation. This intraplanar anisotropy was partially interpreted in terms of the anisotropy of the reciprocal effective masses thermally averaged over several subbands of the valleys [4.29]. The possible anisotropic effect of surface-roughness scattering on (110) wafers has been studied experimentally in this chapter. The mobility data were obtained from two MOS transistors fabricated on the same wafer, with current in the [001] and [$\bar{1}$10] directions, respectively. Substrate doping was $6 \times 10^{15}$ $cm^{-3}$ and $Q_f$ was $1.14 \times 10^{11}$ $cm^{-2}$. The two normalized mobility curves are perfectly matched over the entire range of the vertical electric field (Fig. 4.14a) although the absolute values of mobility will vary in the two surface-current directions (Fig. 4.14b) because of the difference in the effective masses of the inversion-layer carriers. This behavior strongly suggests that surface roughness is isotropic in terms of direction of carrier transport.

The following additional observations should be noted.




a. Normalized mobility




b. Absolute value of mobility

Fig. 4.14. ANISOTROPY OF ELECTRON INVERSION-LAYER MOBILITY VS EFFECTIVE ELECTRIC FIELD ON (11 ) SUBSTRATES.



FCS1689382

FCS1692881

- Except for the difference in the $Q_I$ value in the (110) wafer, the high-field mobility rolloff rate is similar to that of the (111) wafers. This could be related to the oxide growth rate because this growth rate on the (110) surface resembles that of the (111).

- The mobility ratio $[\bar{1}10]/[001]$ is 0.8 (Fig. 4.14b) and is relatively constant over the field region of investigation. This can be compared to 0.87 predicted by the classical theory of diffused scattering with $\tau \approx 2.2 \times 10^{-12}$ sec or to 0.76 as predicted by the classical theory of specular scattering [4.30]. In addition, over the range of impurity doping levels from $9 \times 10^{14}$ to $4 \times 10^{16}$ $cm^{-3}$, this anisotropy ratio was found to be 0.805 ± 0.01.

## 6. Mobility Variation at High Fields on Anisotropically Etched Surfaces

Lower mobility values on etched V-groove surfaces than those of surface devices on the (111) plane have been used in VMOS transistor modeling [4.31], and these values have been attributed to the poorer quality of the V-groove surface. No quantitative data concerning the mobilities of etched surfaces have been reported, however, although such data as a function of gate electric field and chemical etchants are essential in modeling the VMOS and TVMOS devices described in Chapter II.

The test structure in Fig. 4.15 was used to obtain this data. A highly doped $N^+$ (100) starting substrate minimized the drain series resistance, and a boron-doped epitaxial layer was grown for uniform channel doping. The VMOS channel length was determined from spreading-resistance measurements of the spacing between the $N^+$ source diffusion and substrate. Such chemical solutions as KOH and ethylenediamine were chosen as the anisotropic etchants.

Figure 4.16 plots the dependence of effective mobility on gate voltage $V_G - V_T$. Near the threshold voltage where the surface field is low, $\mu_{max}$ is comparable on both the control and etched wafers. The difference in the magnitude of mobility becomes greater as $V_G$ is increased. Based on the following considerations, this mobility difference at high gate voltages can be best explained by surface-roughness scattering.

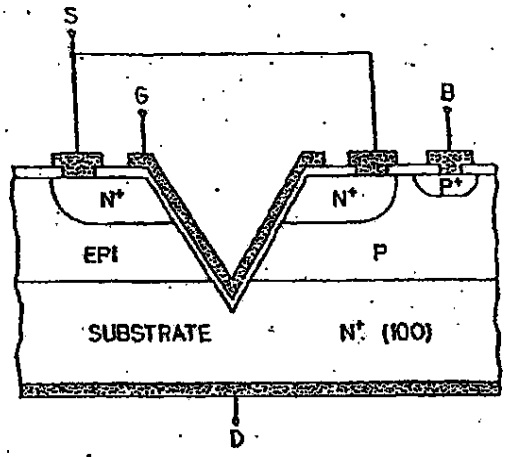



Fig. 4.15. TEST STRUCTURE TO DETERMINE ELECTRON MOBILITY IN INVERSION LAYERS ON ETCHED (111) SURFACES.

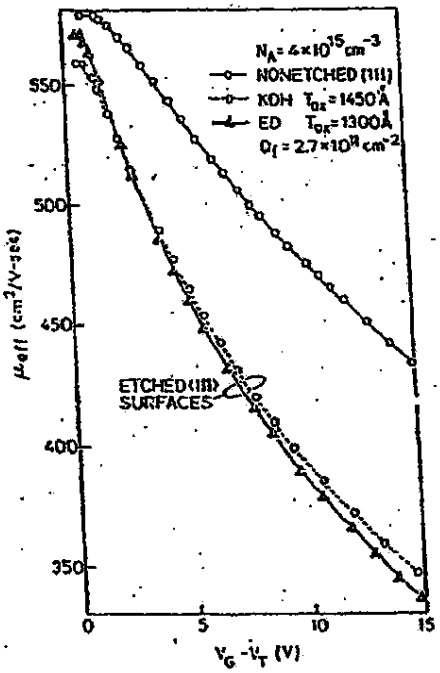



Fig. 4.16. ELECTRON INVERSION-LAYER MOBILITY VS GATE VOLTAGE ON ETCHED AND NONETCHED (111) SURFACES.

- Contrary to silicon-on-sapphire wafers where the epitaxial film quality is poor, mobility reduction resulting from scattering by bulk defects is expected to be small on high-quality epitaxial films grown on single-crystal substrates. The fact that $\mu_{max}$ on the etched surfaces is comparable to that of the control wafer is direct evidence that the effect of bulk defects can be neglected because these defects are best observed at low gate fields and low-inversion electron concentrations near the onset of inversion.

- Properly cleaned wafers after anisotropic etching have essentially the same oxide charge density $Q_f$ as the nonetched wafers. In a separate experiment where C-T measurements were used to monitor any possible contamination resulting from chemical solutions on the planar (111) wafers, the difference in the $Q_f$ values between the etched and nonetched (111) wafers was within the measurement error. The mobility data for $\mu_{max}$ at low fields where coulombic oxide-charge scattering is important also indicate that the scattering caused by charged centers is similar on both the etched and nonetched (111) surfaces.

As discussed in Section B, surface-roughness scattering becomes dominant at high fields and is less significant at low fields, and this should explain the observed behavior. Figure 4.16 indicates that the anisotropically etched silicon surface is rougher than the planar surface and that the difference in surface quality between the two chemical etchants is small.

The mobility variation with gate field can be better observed through normalized mobilities as plotted in Fig. 4.17 where it can be seen that the V-groove surface mobilities degrade faster than those in the control device at high vertical fields. The magnitude of the difference, however, is less than 20 percent even at fields as high as $\approx 3 \times 10^5$ V/cm, which is much less than had been previously reported [4.31]. Another possible implication of the observed mobility degradation caused by stronger surface-roughness scattering is that the reported lower scattering-limited velocity [4.32] compared to nonetched (111) surfaces [4.33] is partially the result of this additional surface roughness.

Figure 4.18 compares the mobilities of etched (111) V-groove surfaces and the (100) surface plane. Over most operating voltage ranges, the planar (100) devices appear to have a 30 percent advantage over the nonplanar devices. This value may change as the ratio of $Q_f$ on the

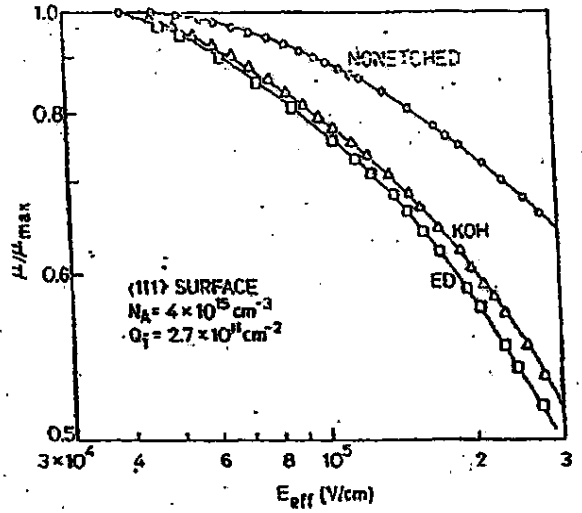



Fig. 4.17. NORMALIZED ELECTRON INVERSION-LAYER MOBILITY VS EFFECTIVE ELECTRIC FIELD ON ETCHED AND NONETCHED (111) SURFACES.

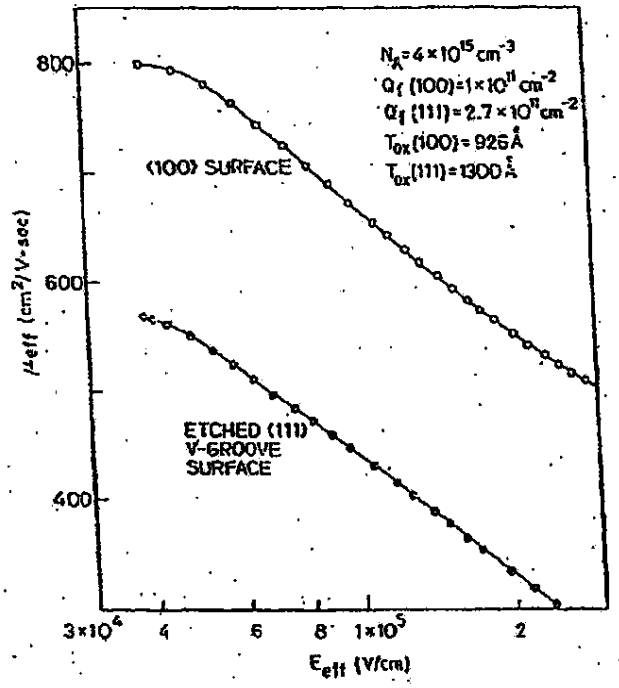



Fig. 4.18. ELECTRON INVERSION-LAYER MOBILITY VS EFFECTIVE ELECTRIC FIELD ON ETCHED (111) AND PLANAR (100) SURFACES.

two surfaces is varied. Both the LDMOS and VDMOS power MOSFETs described in Chapter II can be fabricated on any silicon-crystalline orientation; however, the channel of the VMOS is constrained along an etched <111> surface. Normally, the LDMOS and VDMOS are fabricated on <100> material. Present data indicate that these devices will have a 30 percent advantage in mobility over that of the VMOS, and this results in lower channel resistance and higher device transconductance per unit width in the LDMOS and VDMOS structures.

## E. Electron Accumulation-Layer Mobility

There has been considerably less work on accumulation-layer mobility than on carrier transport in inversion layers although, in recent years, depletion-mode transistors have been used extensively as load devices in MOS integrated circuits and as part of high-voltage MOS structures (see Fig. 2.1). In addition, the dependence of electron accumulation-layer mobility on the processing conditions has not been well characterized. This section describes this dependence on oxide charge density and substrate doping and discusses the observed surface anisotropy on (110) wafers.

In the test structure in Fig. 4.19, $N^-$ epitaxial layers were grown on p-type substrates with (100), (111), and (110) crystal orientations. The epitaxial thickness and resistivity were carefully chosen so that bulk currents would not dominate over the surface current. The structure was isolated by a p-diffusion to reduce fringing currents.

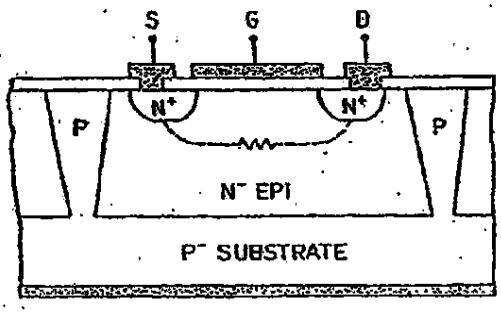



Fig. 4.19. TEST STRUCTURE TO DETERMINE ACCUMULATION-LAYER ELECTRON MOBILITY.

The effective accumulation-layer mobility was determined from dc conductance measurements, with the device operating in the linear region ($V_{DS}$ = 50 mV). When the applied gate potential is such that the device is in the flatband condition (no accumulation or depletion of majority carriers at the silicon surface), the measured drain current becomes the bulk current $I_{DO}$. When the applied potential is greater than flatband voltage $V_{FB}$, it produces an n-type accumulation layer at the surface, and the total drain current increases. This additional current comes from the conduction of the accumulation-layer mobile electrons near the surface. For a given voltage, effective mobility $\mu_{eff}$ can be obtained from the accumulation current defined as

$$I_{accum} = I_D(V_{GS}) - I_{DO}(V_{GS} = V_{FB}) = \left(\frac{W}{L}\right) \mu_{eff} Q_S V_{DS} \qquad (4.15)$$

where

W = MOS channel width

L = channel length

$Q_S$ = total charge induced in the silicon

The flatband voltage was determined from high-frequency C-V measurements of the MOS capacitors. To establish a direct relationship between the accumulation charge $Q_S$ and the corresponding $V_{GS}$, $V_{GS}$ is first related to $Q_S$ [4.34] as

$$V_{GS} = V_{FB} + \psi_S + \frac{Q_S}{C_o} \qquad (4.16)$$

where $C_o$ is the capacitance of the gate oxide and $\psi_S$ is the surface potential. The relationship between $\psi_S$ and $Q_S$ [4.35] is

$$Q_S = \sqrt{\frac{2qN_D\varepsilon_{si}}{\beta}} \left(e^{\beta\psi_S} - \beta\psi_S - 1\right) \qquad (4.17)$$

where

$N_D$ = substrate doping concentration determined by C-V measurements

$\epsilon_{si}$ = permittivity of silicon

$\beta = q/kT$

Effective mobility $\mu_{eff}$ is thus obtained as a function of $V_{GS}$ or $Q_S$.

As with inversion-layer mobility, a plot of accumulation-layer mobility vs average electric field $E_{eff}$ in the accumulation layer may be more meaningful than plotting $\mu$ vs electric field at the silicon surface. The average field is defined as

$$E_{eff} = \frac{\int_0^{X_i} n(x)\, E(x)\, dx}{\int_0^{X_i} n(x)\, dx} = \frac{1/\epsilon_{si} \int_0^{X_i} n(x) \int_c^{X} n(x)\, dx\, dx}{Q_c} = \frac{1}{2}\left(\frac{Q_S}{\epsilon_{si}}\right) \tag{4.18}$$

where $X_i$ is the depth of the accumulation layer from the surface.

The measurements obtained from the (111) and (100) samples with a donor concentration of $N_D = 5 \times 10^{14}$ $cm^{-3}$ are plotted in Fig. 4.20. Electron mobility in the accumulation layer decreases as the surface field is increased. As expected, higher oxide charge densities $Q_f$ result in lower majority-carrier mobility, and the effect of $Q_f$ can be seen over a wide range of gate fields. Unlike inversion-layer mobility, however, a higher $Q_f$ does not have the same degree of influence on the accumulation-layer mobility at low fields because of the difference in spatial distributions between the electron majority and minority carriers (minority carriers are generally distributed closer to the $Si/SiO_2$ interface). As the field is increased, the $Q_f$ influence on both inversion- and accumulation-layer mobilities is observed to be approximately the same. It should be noted that, even at low surface fields, the measured majority-carrier mobility is less than bulk mobility (~80 percent) even for low $Q_f$. Majority-carrier mobility does not depend on the substrate impurity concentration over the range studied from $5 \times 10^{14}$ to $2 \times 10^{15}$ $cm^{-3}$.

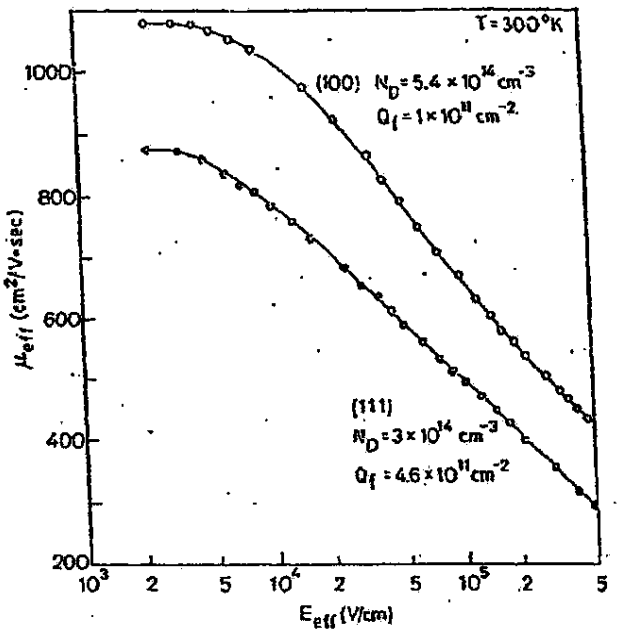



Fig. 4.20. ACCUMULATION-LAYER ELECTRON MOBILITY VS EFFECTIVE ELECTRIC FIELD FOR (111) AND (100) SURFACES.

Electron majority-carrier mobility is compared to inversion-layer mobility in Fig. 4.21 as a function of effective field. The substrate doping level, oxide charge density, and crystalline orientation are the same for both curves. Based on this figure, the following conclusions can be reached.

- Mobility is determined primarily by the surface field rather than carrier concentration. When inversion- and accumulation-layer mobilities are plotted as a function of effective field, the difference is less than 5 percent over most regions. This difference increases, however, when the carrier concentration is used as the horizontal axis.

- The field dependence of accumulation-layer mobility is similar to that of inversion-layer mobility in strong accumulation ($E_{eff} \geq 4 \times 10^4$ V/cm). In the low field (below $3 \times 10^4$ V/cm) where the device is in the sub-threshold region and Eq. (4.2) is no longer applicable,

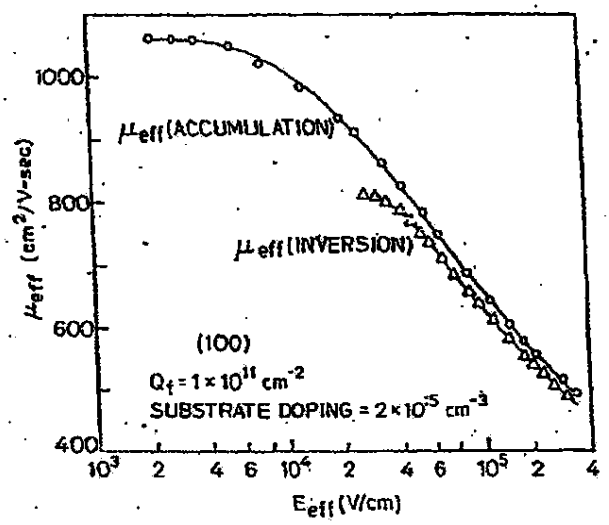



Fig. 4.21. ACCUMULATION- AND INVERSION-LAYER ELECTRON MOBILITIES VS EFFECTIVE FIELD.

inversion-layer mobility is expected to decrease because of the carrier-density fluctuations generated by fixed charge and interface state charges [4.16]. Majority-carrier mobility, however, increases as the field is reduced. No effect of fluctuations is expected in the accumulation layers [4.16] because screening by mobile carriers is greater at high carrier densities. Only on the depletion side of flatband does this carrier density become low enough to allow fluctuations. In addition, accurate measurement of accumulation-layer mobility in weak accumulation is difficult to obtain because of masking by bulk current.

Significant mobility anisotropy in the n-type accumulation layers in the region of strong vertical electric fields at room temperature has been observed in the (110) samples. Figure 4.22 plots the mobilities in two surface directions [$\bar{1}$10] and [001] and their ratio as a function of effective field. The mobility ratio $\mu[\bar{1}10]/\mu[001]$ differs substantially from that of the n-type inversion layers in Fig. 4.14b; anisotropy is small in the low-field region and becomes more significant as the field increases. In very high fields, the mobility ratio approaches that of the inversion layers.

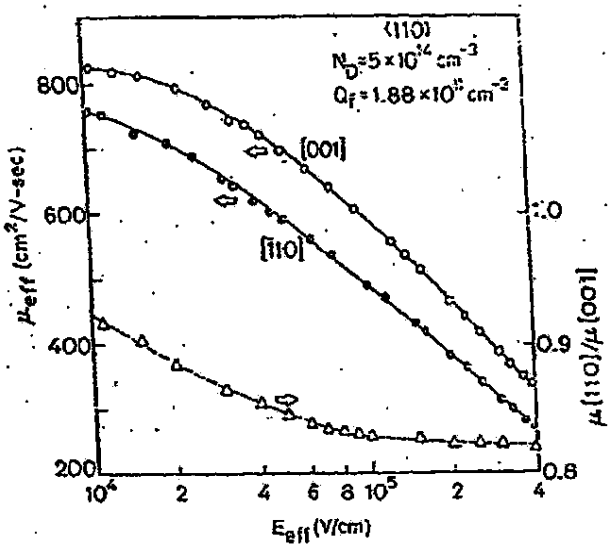



Fig. 4.22. ACCUMULATION-LAYER ELECTRON-MOBILITY ANISOTROPY IN (110) WAFERS.

This behavior in the accumulation layer is similar to the experimental inversion-layer data over a comparable range of surface fields but at lower temperature [4.29]. As in inversion layers, the electric-quantization effects in narrow channels of the accumulation layers have been investigated [4.36,4.37]. These effects are not expected to be significant in low surface fields because of the wide spatial distribution of accumulated carriers, which is quite different from the inverted mobile carriers; they become more important, however, when the carriers are confined in a narrower channel (or potential well) at high fields. Stern and Howard [4.38] reported that electron-mobility anisotropy is expected on the quantized (110) silicon surface and, at room temperature, it may be interpreted in terms of effective-mass anisotropy calculated from the Stern-Howard formula for two-dimensional carriers.

The accumulation-layer mobility data presented in this section are key factors in the calculation of the depletion-mode on-resistance of the power MOSFETs introduced in Chapter II (such as $\mu_n$ in Eq. [2.13]). Because the data available in the literature relate only to a

single-crystal orientation and a single substrate doping concentration, these additional measurements should be useful not only for power MOSFETs but also for small-geometry devices.

F. Summary

This chapter has presented an extensive set of experimental measurements of electron inversion and accumulation-layer mobility which is vital to the successful modeling of the on-resistance and I-V characteristics of the power MOSFETs described in Chapter II. Empirical relationships were developed to predict effective mobility as a function of oxide charge densities and vertical electric fields at low drain voltages. It has been demonstrated that slower oxidation processes [dry vs wet $O_2$ and (100) vs (111)] result in a smoother $Si/SiO_2$ interface and, as a result, slower mobility rolloff under strong vertical fields; this observation also apparently applies to anisotropically etched vs nonetched silicon surfaces, where the nonetched surfaces exhibit a slower high-field mobility rolloff. The implications of these results include opportunities for the optimization of device structures and fabrication technologies as both move toward their ultimate limits.

# Chapter V

## THE RESURF STRUCTURE AND ITS APPLICATIONS TO LDMOS TRANSISTORS

Interest in integrated circuits, especially in MOS transistors operating at high voltage, has increased steadily because they have made possible a very substantial reduction in system costs. Various techniques have been applied to combine high-performance logic circuits with high-voltage transistors, such as the CMOS combined with the DMOS [5.1], silicon-on-sapphire SOS/MOS structures [5.2], and diffused self-aligned (DSA) [5.3] or double-implanted DIMOS [5.4] devices.

To achieve high breakdown voltages, the basic objective is to reduce the surface field so that avalanche occurs only in the bulk. The breakdown mechanisms for vertical power MOSFETs and various junction-termination techniques to minimize the surface effects and to increase the radius of junction curvature have been discussed in Chapter III. This chapter focuses on methods for increasing the breakdown voltage of lateral MOSFETs, particularly the LDMOS.

The LDMOS is attractive for integrated-circuit applications because all three terminals (source, gate, and drain) are located on the topside surface. The drain of the vertical devices, however, is normally the substrate, and a separate diffusion is required, therefore, to make surface contact to the substrate. This will increase either the drain resistance or chip area, which causes the vertical devices to appear less attractive in monolithic circuits.

A recently developed field-shaping technique called RESURF (REduced SURface Field) [5.5] and its principle of operation are described in Section A. Its application to the LDMOS is discussed in Section B and has resulted in transistor breakdown voltages in excess of the planar-junction limit. Variations of the basic RESURF LDMOS are investigated in Section C in an attempt to minimize on-resistance.

### A. RESURF Structure and Principle of Operation

#### 1. Description

The basic RESURF structure (Fig. 5.1) consists of a PN junction formed by a thin N-type epitaxial layer deposited on a P⁻ substrate and